

Research Article

A Comparative Study: Globality versus Locality for Graph Construction in Discriminant Analysis

Bo Yang^{1,2} and Songcan Chen²

¹ College of Computer Science, Inner Mongolia University, Hohhot 010021, China

² Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China

Correspondence should be addressed to Bo Yang; boyangcs@sina.com

Received 23 April 2014; Accepted 27 June 2014; Published 21 July 2014

Academic Editor: Jong Hae Kim

Copyright © 2014 B. Yang and S. Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Local graph based discriminant analysis (DA) algorithms recently have attracted increasing attention to mitigate the limitations of *global* (graph) DA algorithms. However, there are few particular concerns on the following important issues: whether the local construction is better than the global one for intraclass and interclass graphs, which (intraclass or interclass) graph should locally or globally be constructed? and, further how they should be effectively jointed for good discriminant performances. In this paper, pursuing our previous studies on the graph construction and DA, we firstly address the issues involved above, and then by jointly utilizing both the globality and the locality, we develop, respectively, a Globally marginal and Locally compact Discriminant Analysis (GmLcDA) algorithm based on so-introduced global interclass and local intraclass graphs and a Locally marginal and Globally compact Discriminant Analysis (LmGcDA) based on so-introduced local interclass and global intraclass graphs, the purpose of which is not to show how novel the algorithms are but to illustrate the analyses in theory. Further, by comprehensively comparing the Locally marginal and Locally compact DA (LmLcDA) based on locality alone, the Globally marginal and Globally compact Discriminant Analysis (GmGcDA) just based on globality alone, GmLcDA, and LmGcDA, we suggest that the joint of locally constructed intraclass and globally constructed interclass graphs is more discriminant.

1. Introduction

Discriminant analysis (DA) techniques [1] are indispensable in many fields including machine learning, pattern recognition, data compression, scientific visualization, and neural computation. Multiple discriminant analysis (MDA) [2–4] is one of the most popular global DA methods. However, owing to globally constructing both intraclass and interclass graphs, they generally fail to effectively capture underlying local structures in data, for example, many low-dimensional local manifolds of samples residing on the original input space. To mitigate such limitations, plenty of local graph based DA algorithms have been proposed as powerful tools typically including marginal Fisher analysis (MFA) [5] and its variants [6], locality sensitive discriminant analysis (LSDA) [7], LDE [8], and ANMM [9–15]. These algorithms locally construct both intraclass and interclass graphs. However, is the local construction better than the global one for intraclass and interclass graphs? Subsequently, some globally maximizing

and locally minimizing DR algorithms are proposed [16–18]. By contrast, there is no locally marginal and globally compact based DA algorithm to be studied. Several issues need to be addressed, that is, which (intraclass or interclass) graph should locally or globally be constructed? Further, how should they be effectively jointed for good discriminant performances? Up to date, to our knowledge, there are few particular concerns on these issues. So, pursuing our previous studies on the graph construction and DA [19–21], in this paper, we elaborately address the issues involved above. Concretely, firstly, we illustrate the meanings of local compactness, global compactness, local margin, and global margin in DA, as shown in Figure 1; secondly, we formulate globally constructed intraclass and interclass graphs; thirdly, resorting to the relation between the scatter and the structure preservation property of DA based on graph construction, by Proposition 1 and Corollary 2 and Proposition 3 and Corollary 4, we demonstrate that the interclass graph should be globally constructed and the intraclass graph should

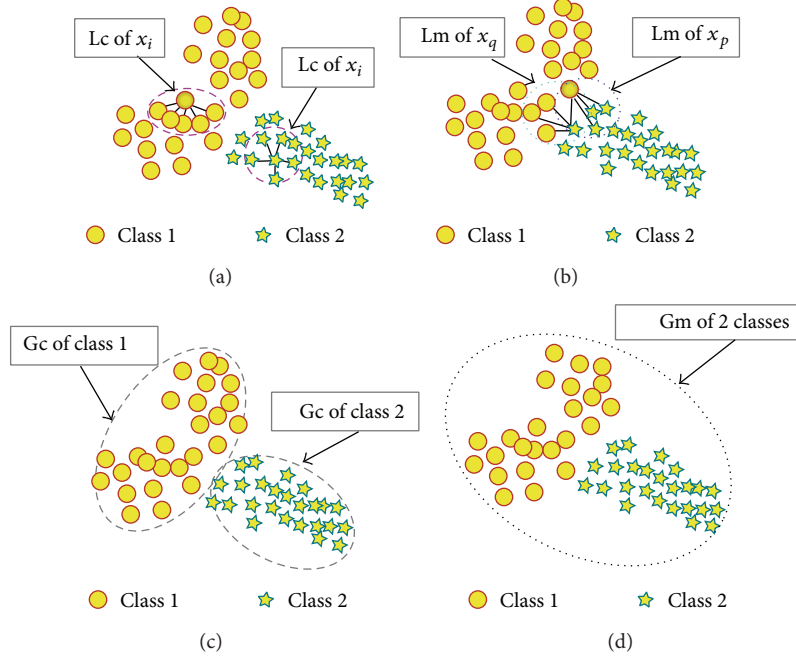


FIGURE 1: Illustration of local and global constructions for intraclass and interclass, where Lc refers to local compactness, Gc to global compactness, Lm to local margin, and Gm to global margin.

be locally constructed; finally, by jointly utilizing both the globality and the locality, we develop two DA algorithms; that is, one is Globally marginal and Locally compact Discriminant Analysis (GmLcDA) algorithm based on so-introduced global interclass and local intraclass graphs, and the other is Locally marginal and Globally compact Discriminant Analysis (LmGcDA) based on so-introduced local interclass and global intraclass graphs. It is worth pointing out that the purpose of developing both DA algorithms is not to show how novel the algorithms are but to illustrate the analyses in theory. Further, we perform experiments to compare GmLcDA, LmGcDA, LmLcDA, and GmGcDA. Concretely, the comparative experiments among GmLcDA, LmGcDA, LmLcDA (MFA and LSDA), and GmGcDA (MDA) are on the toy and real-world datasets. By the comparisons above, we suggest that the joint of locally constructed intraclass and globally constructed interclass graphs is more discriminant. (It is necessary to point out that the two concepts of adjacency matrix and graph are alternatively used in the whole paper since a graph is corresponding to an adjacency matrix.)

The rest of this paper is organized as follows. In Section 2, the graph construction and two typical DA algorithms, MDA and MFA, are briefly reviewed. In Section 3, we firstly indicate the meanings of compactness and margin in DA and introduce the global intraclass graph and interclass graph and then heuristically demonstrate the involved above issues and further develop GmLcDA and LmGcDA and finally compare GmLcDA, LmGcDA, LmLcDA, and GmGcDA. In Section 4, the comparative experiments are performed. Finally, the suggestions and remarks for future work are given in Section 5.

2. Related Works

2.1. Graph Construction. Let $X = \{x_1, \dots, x_n\}$, $x_i \in R^D$ denote a set of n samples; current graph constructions mainly include the two types of k -nearest-neighbor and ε -neighborhood [22]. And the construction of adjacency matrix S is to weight edges of a graph by a similarity function, which mainly refers to the heat kernel and 0-1 two ways [22]. The graph construction of this work focuses on discussing the latter due to its simplicity and generality:

$$S_{ij} = \begin{cases} 1, & \text{if } x_i \text{ is the } k \text{ nearest neighbors of } x_j \\ & \text{or } x_j \text{ is the } k \text{ nearest neighbors of } x_i, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

2.2. Typical DA Algorithms. MDA is deemed an example of GmGcDA here from the viewpoint of graph embedding [5]. Given a dataset of n samples belonging to c classes X with the label $l(x_1), l(x_2), \dots, l(x_n)$, $l(x_i) \in \{1, 2, \dots, c\}$. It seeks the projection directions that maximize the interclass margin and simultaneously minimize the intraclass compactness and thus preserve the global structure in data but fail to discover the local geometric structure in manifold data embedded in the ambient space.

In order to mitigate the limitations of global algorithms, there is increasing interest in graph embedding based DA algorithms. MFA is a typical one, induced from the graph embedding framework for dimensionality reduction [5]. According to the graph embedding framework, MFA constructs a local intraclass graph with the adjacency matrix S to characterize the intraclass compactness and a local

interclass graph with the adjacency matrix S^p to characterize the interclass separability:

$$S_{ij} = \begin{cases} 1, & \text{if } i \in N_{k_1}^+(j) \text{ or } j \in N_{k_1}^+(i), \\ 0, & \text{otherwise,} \end{cases} \quad (2a)$$

$$S_{ij}^p = \begin{cases} 1, & \text{if } (i, j) \in P_{k_2}(c_i) \text{ or } (i, j) \in P_{k_2}(c_j), \\ 0, & \text{otherwise,} \end{cases} \quad (2b)$$

where $N_{k_1}^+(i)$ indicates the index set of the k_1 nearest neighbors of the sample x_i in the same class and $P_{k_2}(c)$ is a set of k_2 nearest sample pairs from different classes.

Here, we call such DA algorithms based on local intra-class and interclass graphs as Locally marginal and Locally compact DA (LmLcDA).

3. Analyzing and Addressing Issues

From the reviews above we have found that motivations of most local graph based DA algorithms are to mitigate the limitations of global algorithms, despite different formulations. However, up to date, there are few particular analyses on whether the local construction is always better than the global one for intraclass and interclass graphs and which (intraclass or interclass) graph should locally or globally be constructed. Further how should they be effectively jointed for good discriminant performances? In this section, in order to further analyze and address these issues, we first illustrate the meanings of local compactness, global compactness, local margin, and global margin in DA, then formally introduce the globally constructed intraclass graph S^{gc} and interclass graph S^{gm} , elaborately analyze and address the issues involved above, and finally develop Globally marginal and Locally compact Discriminant Analysis (GmLcDA) and Locally marginal and Globally compact Discriminant Analysis (LmGcDA).

3.1. Meanings of Compactness and Margin in DA. Now we firstly illustrate the meanings of local compactness, global compactness, local margin, and global margin in DA, respectively, which are all shown in Figure 1. Figure 1(a) shows the structures of local compactness of x_i and x_j , where x_i is locally linked with the five dots within the same class and x_j with pentacles, and such structures of local compactness are encompassed by the two pink dash-line ellipses for the sake of clearer display. Meanwhile, Figure 1(b) shows the structures of local margin of x_p and x_q , where x_p is locally linked with the four pentacles from the different classes, x_q with dots, and such structures of local margin are encompassed, respectively, by the blue and cyan dot-line ellipses. By contrast, Figure 1(c) shows the structures of global compactness of classes 1 and 2, while Figure 1(d) shows the structures of global margin of both classes. And it should be noted that the gray dash-line and dot-line ellipses do not denote a cluster but all points linked within them.

3.2. Globally Constructed Intraclass and Interclass Graphs. The global intraclass graph S^{gc} is formulated as follows:

$$S_{ij}^{\text{gc}} = \begin{cases} 1, & \text{if } l(x_i) \text{ and } l(x_j) \text{ belong to the same class,} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

and the global interclass graph S^{gm} as follows:

$$S_{ij}^{\text{gm}} = \begin{cases} 1, & \text{if } l(x_i) \text{ and } l(x_j) \text{ belong to different classes,} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

From the formulation of S^{gc} in (3) and S^{gm} in (4), it can be seen that globally constructed intraclass and interclass graphs are parameter-free. In order to compare them with locally constructed graphs, their neighbor parameters may, respectively, be viewed as $\sum_{i=1}^c n_i^2$ and $n^2 - \sum_{i=1}^c n_i^2$ for a dataset with c classes and n samples (n_i samples per class), where $\sum_{i=1}^c n_i^2$ denotes the maximum number of neighbor sample pairs intraclass.

3.3. Heuristic Demonstration of Issues. In this subsection, resorting to the relation between the scatter and the structure preservation property of DA based on graph construction, by Proposition 1 and Corollary 2 and Proposition 3 and Corollary 4, we demonstrate that the interclass graph should be globally constructed and the intraclass graph should be locally constructed. Those inequalities in the two propositions and corollaries demonstrate the scatter discrepancies between locally constructed graph and globally constructed graph in the input space. And on the other hand, there is a geometry structure preservation hypothesis; that is, the intraclass graph in DA can preserve the compact structures of the input space into the embedded space, while the interclass graph preserves the margin structures of the input space into the embedded space. Under such hypothesis, those scatter inequalities heuristically demonstrate to some extent that the intraclass graph should be locally constructed and the interclass graph should be globally constructed.

Proposition 1. For a locally constructed intraclass graph corresponding to the adjacency matrix S^t with parameter t , then the intraclass scatters corresponding to S^{t_1} and S^{t_2} are

$$\sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{t_1} < \sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{t_2}, \quad \text{if } t_1 < t_2, \quad (5)$$

where the parameter t stands for the nearest neighbors of the sample x_i or x_j in the same class, as k_1 defined in (2a), while the parameters t_1 and t_2 denote that t in S^t takes t_1 and t_2 , respectively, for S^{t_1} and S^{t_2} .

Proof. According to the definition of the adjacency matrix S^t , its element S_{ij}^t takes 1 or 0. What take value as 1 in S^{t_1} is less than that in S^{t_2} if $t_1 < t_2$; and $\|x_i - x_j\|^2 > 0$ when $i \neq j$. Hence, for $t_1 < t_2$, the intraclass scatters are $\sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{t_1} < \sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{t_2}$. The proof is completed. \square

Corollary 2. $\sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{k_1} \leq \sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{g_c}$.

From Proposition 1 it can be seen that Corollary 2 is clear since the parameter $k_1 \leq \sum_{i=1}^c n_i^2$.

From Proposition 1 and Corollary 2 it can be shown that the intraclass scatter corresponding to locally constructed graph in the input space is not larger than that corresponding to globally constructed graph. And, it is well known that the intraclass graph for DA algorithms aims to preserve the local compactness structures of intraclass samples in the input space into the embedded space, as shown by Figure 1(a) not Figure 1(c). So, according to the geometry structure preservation property of graph embedding DA algorithms, small intraclass scatter in the input space is often also small in the embedding space; in other words, large intraclass scatter in the input space is often also large in the embedding space; then the compactness of intraclass samples corresponding to locally constructed graph often can be preserved in the low-dimensional space. Thus intraclass graph often should locally but not globally be constructed, which is consistent with the statement in [23] that, empirically, small neighbor parameter tends to perform better.

Proposition 3. For a locally constructed interclass graph corresponding to the adjacency matrix $S^{p,g}$ with parameter g , then the interclass scatters corresponding to S^{p,g_1} and S^{p,g_2} are

$$\sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{p,g_2} > \sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{p,g_1}, \quad \text{if } g_2 > g_1, \quad (6)$$

where the parameter g stands for the nearest sample pairs from different classes, as k_2 defined in (2b), while the parameters g_1 and g_2 denote that g in $S^{p,g}$ takes g_1 and g_2 , respectively, for S^{p,g_1} and S^{p,g_2} .

The proof is similar to Proposition 1 and thus omitted.

Corollary 4. $\sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{g_m} \geq \sum_{i,j=1}^n \|x_i - x_j\|^2 S_{ij}^{p,k_2}$.

From Proposition 3 it can be seen that Corollary 4 is clear since $(n^2 - \sum_{i=1}^c n_i^2) \geq k_2$.

The interclass graph for DA algorithms aims to effectively separate samples from different classes: thus the local construction of interclass graph is not quite reasonable on several facets. (1) It is expected that the samples from different classes are separated as effectively as possible, as shown in Figure 1(d) not Figure 1(b); that is, the interclass scatter is as large as possible, while the interclass scatter corresponding to local graph in the input space is not larger than that corresponding to global graph, as demonstrated in Proposition 3 and Corollary 4. Considering the properties of structure preservation for graph embedding DA algorithms, the local construction for interclass graph seems not desirable. (2) The exhaustive search for the neighbor parameter (such as k_2 in MFA) in high-dimensional input space is unavoidable and very expensive for a good performance. Moreover, if the parameter is unsuitably set, then the supervised information cannot be sufficiently effectively utilized, and further it seems

not so easy to confirm which parameter value on earth is most suitable to a task. By contrast, the global construction of interclass graph does not need exhaustive search for the neighbor parameter anymore because the neighbor parameter is adaptive to different datasets. Furthermore, all available supervised information as prior knowledge can be sufficiently utilized in maximizing the interclass margin. So, interclass graph often should globally but not locally be constructed, for good discriminant performances.

From the discussions above, it can be expected that the joint of local construction for intraclass and global one for interclass is more discriminant for good performances. Next, in order to illustrate the analyses above, we develop two DA algorithms of GmLcDA and LmGcDA.

3.4. GmLcDA and LmGcDA. In this subsection, two DA algorithms are developed: one is the Globally marginal and Locally compact Discriminant Analysis (GmLcDA) by jointly utilizing the global interclass graph S^{g_m} in (4) and the local intraclass graph S^{l_c} in (2a) and the other is Locally marginal and Globally compact Discriminant Analysis (LmGcDA) by jointly utilizing the local interclass graph S^p in (2b) and the global intraclass graph S^{g_c} in (3).

In order to develop GmLcDA, the adjacency matrix S in (2a) is rewritten as S^{l_c} corresponding to the local intraclass graph:

$$S_{ij}^{l_c} = \begin{cases} 1, & \text{if } x_i \in N_{k_c}(x_j) \text{ or } x_j \in N_{k_c}(x_i), \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where $N_{k_c}(x_i)$ denotes the set of k_c nearest neighbors of sample x_i within the same class. In fact, S^{l_c} in (7) is consistent with S in (2a). However, since GmLcDA only involves one neighbor parameter for the local intraclass graph (the global interclass graph is parameter-free), the formulation with parameter k_c in (7) is used rather than the one with parameter k_1 in (2a) for a clearer understanding.

Now compute the interclass scatter B^{g_m} and intraclass scatter A^{l_c} ,

$$\begin{aligned} B^{g_m} &= \sum_{i,j} (y_i - y_j)^2 S_{ij}^{g_m} \\ &= \sum_{i,j} \|w^T x_i - w^T x_j\|^2 S_{ij}^{g_m} \\ &= 2w^T \left[\sum_i x_i D_{ii}^{g_m} x_i^T - \sum_{i,j} x_i S_{ij}^{g_m} x_j^T \right] w \\ &= 2w^T X (D^{g_m} - S^{g_m}) X^T w \\ &= 2w^T X L^{g_m} X^T w, \end{aligned} \quad (8)$$

where D^{g_m} is a diagonal matrix and its entries are column (or row, since S^{g_m} is symmetric) sum of S^{g_m} , $D_{ii}^{g_m} = \sum_j S_{ij}^{g_m}$, and the corresponding Laplacian matrix $L^{g_m} = D^{g_m} - S^{g_m}$ is symmetric and positive semidefinite matrices as well.

Likely, A^{lc} is obtained by replacing S^{gm} in (8) with S^{lc} and correspondingly we obtain D^{lc} and L^{lc} .

Further, the objective criterion of GmLcDA can be formulated by jointly maximizing B^{gm} and minimizing A^{lc} to simultaneously preserve the local compactness structures of intraclass data and the margins between different-class data. Consider

$$\begin{aligned} w^{GmLc} &= \arg \max_w \frac{B^{gm}}{A^{lc}} = \arg \max_w \frac{w^T X (D^{gm} - S^{gm}) X^T w}{w^T X (D^{lc} - S^{lc}) X^T w} \\ &= \arg \max_w \frac{w^T X L^{gm} X^T w}{w^T X L^{lc} X^T w} \end{aligned} \quad (9)$$

which can be solved by the generalized eigen-decomposition $XL^{gm}X^T w = \lambda XL^{lc}X^T w$ [24].

Likely, for LmGcDA, the local interclass adjacency matrix S^p is rewritten as S^{lm} corresponding to the local interclass graph:

$$S_{ij}^{lm} = \begin{cases} 1, & \text{if } (i, j) \in P_{km}(c_i) \text{ or } (i, j) \in P_{km}(c_j), \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where $P_{km}(c)$ is a set of km nearest sample pairs from different classes. Further S^{lm} in (10) and S^{gc} in (3) are jointed to form the objective criterion of LmGcDA as follows:

$$\begin{aligned} w^{LmGc} &= \arg \max_w \frac{B^{lm}}{A^{gc}} = \arg \max_w \frac{w^T X (D^{lm} - S^{lm}) X^T w}{w^T X (D^{gc} - S^{gc}) X^T w} \\ &= \arg \max_w \frac{w^T X L^{lm} X^T w}{w^T X L^{gc} X^T w} \end{aligned} \quad (11)$$

which can be solved by the generalized eigen-decomposition $XL^{lm}X^T w = \lambda XL^{gc}X^T w$ [24].

3.5. Comparison among GmLcDA, LmGcDA, LmLcDA, and GmGcDA. GmLcDA jointly constructs global interclass graph, as shown in Figure 1(d), and local intraclass graph, as shown in Figure 1(a). Such graph construction leads to larger margin between classes and larger compactness within the same class, so it can preserve the geometry structure and is consistent to Proposition 1 and Corollary 2; Proposition 3 and Corollary 4. As a result, it more likely results in good discriminant performances, as shown in the experiments in Section 4.

LmGcDA exactly shows an opposite effect. That is, it is difficult for a locally constructed interclass graph (shown in Figure 1(b)) and a globally constructed intraclass graph (shown in Figure 1(c)) to effectively preserve geometry structure of input samples, in fact, which leads to the worse performances in real-world tasks, as shown by the experiments in Section 4.

By contrast, LmLcDA constructs locally both intra- and interclass graphs, as shown in Figures 1(b) and 1(a), respectively. It is almost well known that the local parameter settings

of graph construction are relatively intractable, especially for interclass. Accordingly, for good discriminant performance, such construction of graph needs more domain knowledge and experience.

Different from the three above, the graph construction of GmGcDA only adopts the global way, for both intraclass and interclass. For such construction, its accomplishment is easy and stable for different domains.

4. Experiments

In this section, to further illustrate and support the analyses in the above theory, we compare LmLcDA, GmLcDA, LmGcDA, and GmGcDA by performing the experiments on the toy and the real-world datasets; the latter includes UCI [25], face recognition, and object categorization. The nearest neighbor classifier (1NN) [26] is followed after these DA algorithms to evaluate their classification performances. The ridge regularization [27] is adopted for all compared DA algorithms. That is, all the compared DA algorithms are derived from a regularized objective of the same form as $w^* = \arg \max(w^T B w) / (w^T A w + \alpha^* I)$ with parameter $\alpha = 0.1$, where B and A , respectively, respond to inter- and intraclass scatters.

4.1. A Toy Example. Here, the toy example illustrates that both intraclass and interclass graphs locally constructed show less discriminant projections, implying that the global construction appears more necessary and thus suggesting that a globally constructed interclass graph and a locally constructed intraclass graph can be desirable to obtain more discriminant performance.

For LmLcDA, such as MFA, due to the injection of the locality into both intraclass and interclass graphs, they may fail to obtain discriminant projections for some tasks that the global structure needs to consider. Moreover, there is not any guidance on the selection of the neighbor parameter for the interclass graph. When the neighbor parameters for constructing graph are inappropriately set, their projection results may be more unfavorable. For example, for a linearly separable problem shown in Figure 2(a), for the DA algorithms that adopt at least one global graph, GmGcDA (MDA), GmLcDA with parameter $kc = 3$ and LmGcDA with $km = 5$ can effectively separate the two-class samples, as shown in Figures 2(c)–2(e). On the contrary, from Figure 2(b), we can observe that, in the reduced one-dimensional subspace of MFA (k_1 is empirically set to 3 and different k_2 's), the two-class samples overlap to different extents. Moreover, no laws can be found for the selection of k_2 . We only can observe that when $k_2 = 15$, the separation of two class samples is relatively satisfactory.

Further while the input points are added, as shown in Figure 2(f), it is changed that both the margin between the two classes and the compactness within Class 2. And it is more difficult to partition the two-class samples in the one-dimensional projection space of LmGcDA than in the one of GmLcDA, as shown in Figures 2(g) and 2(h).

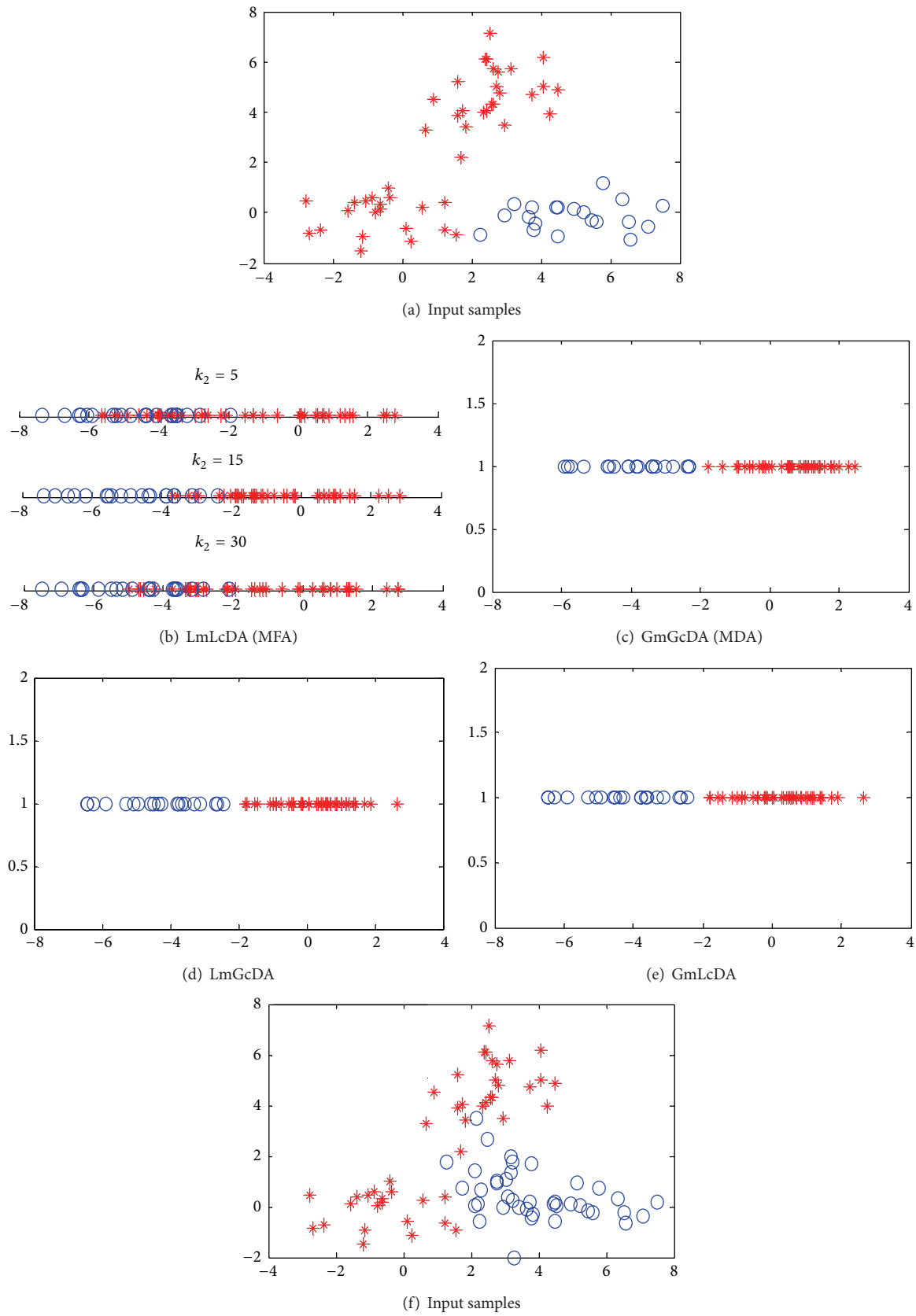


FIGURE 2: Continued.

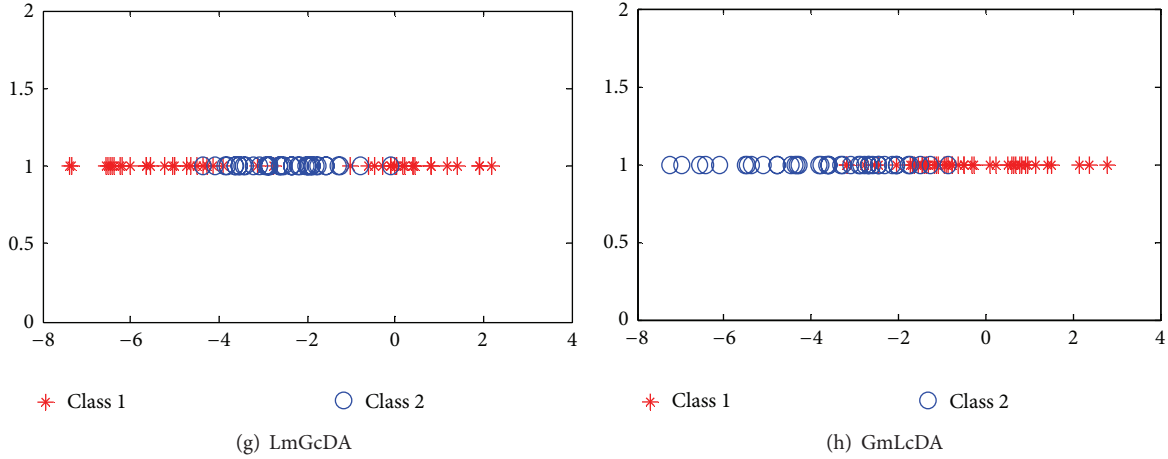


FIGURE 2: A toy example.

4.2. Real-World Datasets. On these real-world datasets, we compare GmLcDA, LmGcDA, two LmLcDAs (MFA and LSDA), and GmGcDA (MDA). In order to effectively evaluate the various algorithms, their model parameters are searched from a large candidate range and correspondingly the best results are reported. To address the singularity of MDA, here the inverse of matrix S_w is replaced by the pseudoinverse [24].

4.2.1. UCI Datasets. We select the 5 two-class UCI datasets, whose descriptions are shown in column 1 of Table 1, and the classification results of INN on the original data are reported.

For each dataset, samples are randomly divided into training set and testing set, which, respectively, contain half of the samples. The random division is performed 30 times and the average accuracies for each algorithm are tabulated in Table 1. The neighbor parameters kc for GmLcDA, km for LmGcDA, k_1 for MFA, and k for LSDA are searched from 2 to half of the minimum of each class sample with the increment of 5. The parameter k_2 for MFA is searched from 20 to the sum (Notice that the maximum of k_2 for MFA, i.e., the sum of training samples (n), is smaller than the so-called parameter of global interclass graph for GmLcDA ($n^2 - \sum_{i=1}^c n_i^2$)) of training samples with the increment of 20.

From the results shown in Table 1, we can observe the following.

- (i) GmLcDA produces the optimal accuracies on the 4 datasets except for Sonar, which clearly outperforms LmGcDA, MDA, MFA, LSDA, and INN with the unreduced data. The accuracy of GmLcDA only is 0.0048, lower than LSDA on Sonar. Moreover, the optimal reduced dimensions of GmLcDA (except for Sonar) are lower than those of MFA and LSDA, which clearly is helpful for efficient testing.
- (ii) LmGcDA is the worst on all datasets except for Crx, on which it is only better than MDA. Such results clearly testify that the locality for interclass graph and globality for intraclass graph are not alternatives.

- (iii) Both MFA and LSDA achieve almost similar results since they adopt local intraclass and interclass graphs to preserve the local geometry of data. However, the optimal reduced dimensions of MFA are higher than those of LSDA overall.
- (iv) The performances of MDA are relatively inferior, especially on Crx and Spectf and even worse than INN on the unreduced data. Moreover, the standard deviations of MDA on all datasets seem larger than those of the other algorithms, which can be attributed to the factor that its reduced dimension can only be limited to one dimension for two-class datasets.
- (v) GmLcDA, LmLcDA, and GmGcDA almost all outperform INN on all datasets (except for GmGcDA on Crx, Sonar, and Spectf) with lower reduced-dimensions. By contrast, LmGcDA is worse than INN on all datasets.

4.2.2. Face Recognition. It is well known that face recognition is a very important task in pattern recognition and machine learning. In this subsection, two benchmark datasets, Yale and ORL, are used to evaluate the uses of local and global graphs for DA algorithms on the face image recognition.

Data Description. The Yale dataset contains 165 face images of 15 individuals ($c = 15$), 11 images per individual, and these 11 images were gotten at different facial expressions or configuration. Figure 3(a) is 11-sample face images of one person in Yale dataset. The ORL dataset consists of 400 face images for 40 distinct subjects ($c = 40$), 10 images per subject. These images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). Figure 3(b) is 10 sample face images of one person in ORL dataset. The size of each image in Yale and ORL is 112×92 and 100×100 pixels, with 256 grey levels per pixel.

Experimental Settings. Considering the very high dimension of the face images, in our experiment, PCA [28, 29] is firstly

TABLE 1: Accuracy% \pm standard deviation (optimal reduced dimensions), where the reduced highest dimension of MDA is number of class subtract one, here equals one.

Datasets (number of sample, number of attr.)	GmLcDA	LmGcDA	LmLcDA (MFA)	LmLcDA (LSDA)	GmGcDA (MDA)	INN
Crx (666, 6)	70.63 \pm 1.22 (2)	66.42 \pm 5.67 (6)	69.76 \pm 0.79 (6)	68.71 \pm 1.74 (6)	59.58 \pm 7.47	67.29 \pm 1.33
Sonar (208, 60)	85.15 \pm 2.63 (42)	62.68 \pm 0.65 (10)	83.88 \pm 1.66 (24)	85.63 \pm 3.03 (34)	69.61 \pm 5.81	83.11 \pm 3.08
Spectf (267, 44)	77.97 \pm 2.88 (6)	67.52 \pm 3.92 (10)	76.47 \pm 1.77 (24)	75.94 \pm 1.88 (4)	68.72 \pm 5.13	71.13 \pm 3.90
Water (116, 38)	95.44 \pm 2.06 (2)	92.54 \pm 0.75 (4)	94.91 \pm 3.83 (8)	92.11 \pm 4.40 (2)	90.88 \pm 4.81	87.90 \pm 3.55
Wdbc (569, 30)	96.23 \pm 1.28 (6)	80.51 \pm 2.58 (10)	95.53 \pm 0.86 (30)	95.78 \pm 0.79 (10)	94.97 \pm 1.38	91.69 \pm 1.13

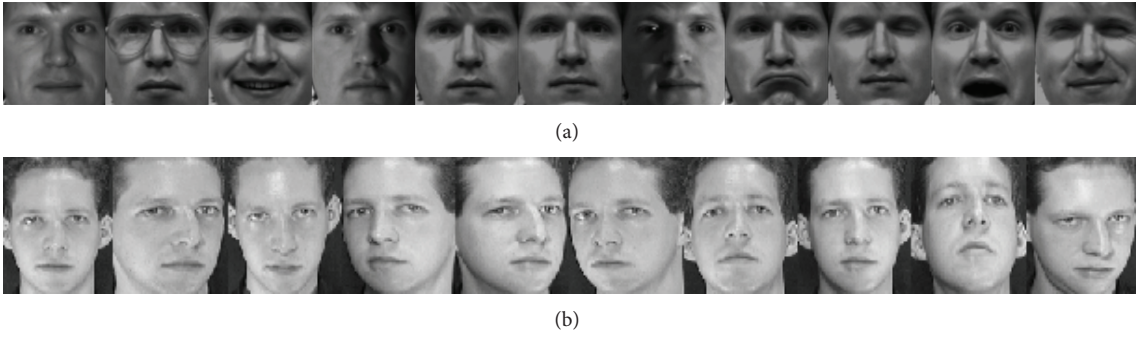


FIGURE 3: Face sample images of one person. (a) Yale, (b) ORL.

adopted to identify a low-dimensional subspace. Here, 99% energy is kept. Then the various DA algorithms are performed in the obtained PCA subspace. In order to show effects of the various algorithms on training sets with different numbers of samples, each face dataset is partitioned into the different gallery and probe sets where G_m/P_n indicates that m images for each person are randomly selected for training and the remaining n images for testing. For each dataset, 30 random splits with G_m/P_n are generated and the average results of the 30 classification accuracies are reported. The neighbor parameters kc for GmLcDA and k_1 for MFA are searched from $\{2, 3, \dots, G_m - 1\}$, km for LmGcDA and k_2 for MFA from $\{5, 10, \dots, c^*G_m\}$ on Yale, from $\{20, 40, \dots, c^*G_m\}$ on ORL, and k for LSDA from $\{5, 10, \dots, c^*G_m/2\}$. The optimal results of the several algorithms on the two datasets are listed in Table 2.

From the results of Table 2, we can observe the following.

- (i) With the increase of gallery samples, that is, $G_m = 3$ to 7 (5) on Yale (ORL), the performances of these algorithms increase to different degrees.
- (ii) For the two datasets with different G_m/P_n divisions, GmLcDA always obtains better accuracies than the other algorithms, which shows effectiveness and feasibility for face recognition task. These relatively excellent results of GmLcDA may be ascribed to the

joint injection of the global margin interclass and local compactness intraclass.

- (iii) MFA and LSDA produce some worse results than MDA, such as on Yale with G_4/P_7 and G_5/P_6 for MFA, G_6/P_5 for LSDA, and on ORL with G_5/P_5 for MFA. Moreover, MFA is relatively worse. A possible reason is that only the local geometry of data for intraclass and interclass limits their performances, especially MFA, whose k_2 marginal points possibly limit its generalization ability to a certain degree.
- (iv) On Yale, the unsupervised PCA is worse than the other algorithms. However, it is worth noting that PCA produces better accuracies than MDA, LSDA, and MFA on ORL. That may be ascribed to such a factor that, for ORL, the margin of different-class samples is larger in the PCA subspace with 99% energy. Besides, on Yale, the standard deviations of MDA and PCA are larger than those of the other algorithms.

4.2.3. Object Categorization. A classical problem in computer vision and pattern recognition is to classify a set of objects into a group of known categories. Here, we use the popular benchmark dataset Coil20. The dataset consists of gray-scale images of 20 objects and 72 images for each object with pose

TABLE 2: Accuracy% \pm standard deviation (optimal reduced dimensions), where the reduced highest dimension of MDA is number of class subtract one, here equals one.

Dataset	G_m/P_n	GmLcDA	LmGcDA	LmLcDA (MFA)	LmLcDA (LSDA)	GmGcDA (MDA)	PCA
Yale	G_3/P_8	84.58 \pm 2.89 (14)	69.04 \pm 6.18 (10)	77.25 \pm 3.73 (16)	79.42 \pm 5.08 (18)	72.17 \pm 3.89	64.25 \pm 3.23
	G_4/P_7	88.29 \pm 4.09 (14)	73.95 \pm 6.52 (10)	82.10 \pm 4.18 (16)	85.62 \pm 4.27 (18)	85.52 \pm 5.25	67.43 \pm 2.83
	G_5/P_6	89.78 \pm 3.13 (16)	76.67 \pm 6.23 (10)	86.33 \pm 3.14 (24)	88.11 \pm 3.52 (20)	89.22 \pm 2.67	68.22 \pm 3.52
	G_6/P_5	91.33 \pm 2.76 (14)	82.80 \pm 4.70 (10)	89.47 \pm 2.22 (18)	89.73 \pm 3.50 (24)	90.67 \pm 2.81	70.13 \pm 4.45
	G_7/P_4	93.17 \pm 3.55 (28)	84.67 \pm 5.06 (10)	93.00 \pm 3.41 (20)	92.00 \pm 3.67 (20)	92.33 \pm 4.10	73.00 \pm 6.56
	G_3/P_7	90.46 \pm 2.36 (38)	86.45 \pm 2.60 (24)	88.68 \pm 1.95 (28)	87.43 \pm 2.03 (39)	84.89 \pm 2.47	88.29 \pm 2.79
ORL	G_4/P_6	93.83 \pm 1.16 (43)	87.00 \pm 2.29 (18)	90.88 \pm 2.49 (18)	90.92 \pm 1.19 (40)	90.67 \pm 1.42	91.83 \pm 1.33
	G_5/P_5	95.25 \pm 1.30 (78)	85.45 \pm 3.72 (16)	90.10 \pm 3.33 (23)	93.65 \pm 2.01 (40)	93.45 \pm 1.46	94.65 \pm 1.18



FIGURE 4: Sample images of 20 objects from Coil20.

intervals of 5° , the size of each image being 32×32 pixels, as shown in Figure 4.

Experimental Settings. Similar to the face recognition experiment, the dataset is partitioned into the different gallery and probe sets with 4 groups. For each group G_m/P_n , 20 random splits are generated and the average results of the 20 classification accuracies are reported. For the first three groups with relatively small gallery sets ($G_m/P_n = G_6/P_{66}$, G_9/P_{63} and G_{18}/P_{54}), the neighbor parameters kc for GmLcDA and k_1 for MFA are searched from $\{2, 4, \dots, G_m - 1\}$, km for LmGcDA and k_2 for MFA from $\{20, 40, \dots, c^*G_m\}$, and k for LSDA from $\{10, 30, \dots, c^*G_m - 2\}$. For relatively

large gallery samples with G_{36}/P_{36} , kc and k_1 are searched from $\{2, 8, \dots, G_m - 1\}$, km and k_2 from $\{40, 120, \dots, c^*G_m\}$, and k from $\{10, 90, \dots, c^*G_m - 2\}$. The accuracies with respect to different reduced dimensions are displayed in Figure 5.

From Figure 5 we can clearly see the following.

- (i) Figures 5(a)–5(d) show that, with the increase of gallery samples, the performances of these algorithms are improved to different extents. Likewise, with the gradual increase of reduced dimension, their accuracies correspondingly increase. However, with the further increase of reduced dimension, their performances (except for MDA whose reduced dimension is at most $c - 1$) decrease gradually. This result is

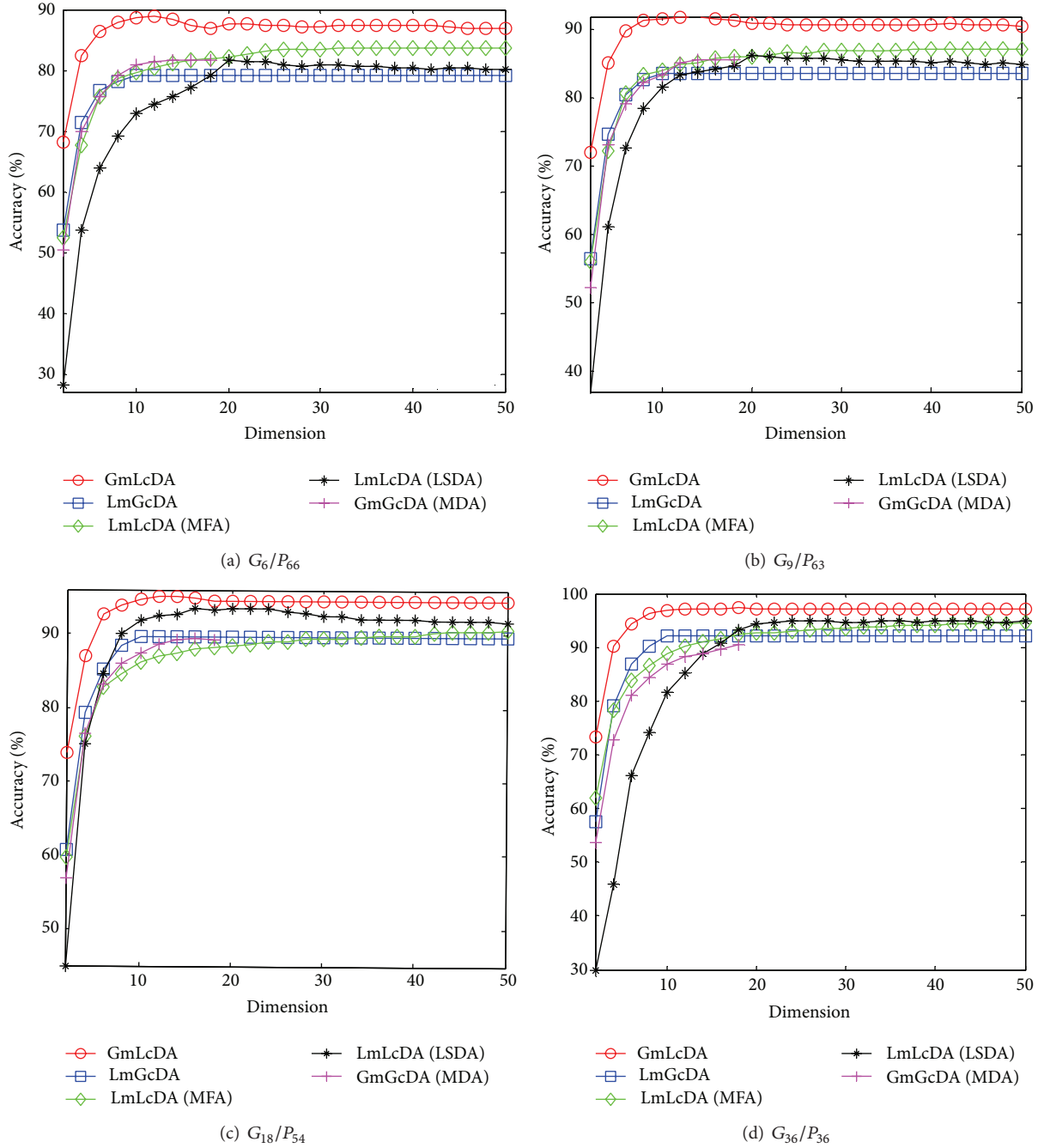


FIGURE 5: Accuracies with respect to the reduced dimensions on Coil20.

especially evident for small gallery samples with $G_m = 6$, which undoubtedly is an example of “curse of dimensionality” [2].

- (ii) For the four groups G_m/P_n from small G_6 to relative large G_{36} , GmLcDA outperforms all MFA, LSDA, and MDA with respect to different reduced dimensions. In particular, when the reduced dimension is less than 10, its superiority is more obvious. For example, when the reduced dimension is 2, it is over 30% better than the worst LSDA. Moreover, the reduced dimension of

GmLcDA with respect to its best accuracy is lower than that of the other algorithms. We can observe that the accuracies of GmLcDA have achieved the best value when the reduced dimension is less than or equal to 20, and afterward they tend to maintain stable.

- (iii) The performances of LSDA and MFA do not show very remarkable distinction except for $G_m = 18$. Specifically, for the first two groups, MFA is more predominant than LSDA overall, as shown in

Figures 5(a) and 5(b). In contrast, with the increase of the gallery sample number, LSDA exceeds MFA in performance, as shown in Figures 5(c) and 5(d). We basically claim that LSDA is not suitable to small training samples. Moreover, we can very clearly observe that the performances of LSDA are always the worst when the reduced dimension is relatively low.

- (iv) When the reduced dimensions are less than the reduced range ($c - 1$) of MDA, the accuracies of MDA are almost parallel to those of MFA and LSDA. However, since the reduced dimension is only limited to the maximum of $c - 1$, its performances are inferior to the other three algorithms overall.

5. Conclusion and Future Work

In this paper, we elaborately address some important issues in DA based on graph construction. And in order to illustrate and support the analyses in theory, by jointly utilizing both the globality and the locality, we develop GmLcDA algorithm based on the global interclass and local intraclass graphs and LmGcDA based on the local interclass and global intraclass graphs. Further, by comprehensively comparing LmLcDA (MFA, LSDA), GmLcDA, LmGcDA, and GmGcDA (MDA) on toy and real-world datasets, we suggest that the joint of locally constructed intraclass and globally constructed interclass graphs is more discriminant.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by NSFC (61170151, 61363051), JS-QingLan Project, Program of Higher-Level Talents of Inner Mongolia University (115118), Postdoctoral Science Foundation funded Project of Inner Mongolia University (30105-135113), General Financial Grant from China Postdoctoral Science Foundation (2013M540217).

References

- [1] K. Bunte, M. Biehl, and B. Hammer, "A general framework for dimensionality-reducing data visualization mapping," *Neural Computation*, vol. 24, no. 3, pp. 771–804, 2012.
- [2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2001.
- [3] J. H. Oh and N. Kwak, "Generalization of linear discriminant analysis using L_p -norm," *Pattern Recognition Letters*, vol. 34, no. 6, pp. 679–685, 2013.
- [4] J. Liu, F. Zhao, and Y. Liu, "Learning kernel parameters for kernel Fisher discriminant analysis," *Pattern Recognition Letters*, vol. 34, no. 9, pp. 1026–1031, 2013.
- [5] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.
- [6] D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang, "Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2811–2821, 2007.
- [7] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI '07)*, pp. 708–713, January 2007.
- [8] H. Chen, H. Chang, and T. Liu, "Local discriminant embedding and its variants," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 846–853, June 2005.
- [9] F. Wang and C. Zhang, "Feature extraction by maximizing the average neighborhood margin," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.
- [10] M. H. Rohban and H. R. Rabiee, "Supervised neighborhood graph construction for semi-supervised classification," *Pattern Recognition*, vol. 45, no. 4, pp. 1363–1372, 2012.
- [11] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV '07)*, pp. 1–7, Rio de Janeiro, Brazil, October 2007.
- [12] M. M. Luqman, J. Ramel, J. Lladós, and T. Brouard, "Fuzzy multilevel graph embedding," *Pattern Recognition*, vol. 46, no. 2, pp. 551–565, 2013.
- [13] Y. Song, F. Nie, C. Zhang, and S. Xiang, "A unified framework for semi-supervised dimensionality reduction," *Pattern Recognition*, vol. 41, no. 9, pp. 2789–2799, 2008.
- [14] H. Bunke and K. Riesen, "Recent advances in graph-based pattern recognition with applications in document analysis," *Pattern Recognition*, vol. 44, no. 5, pp. 1057–1067, 2011.
- [15] S. Park and S. Choi, "Max-margin embedding for multi-label learning," *Pattern Recognition Letters*, vol. 34, no. 3, pp. 292–298, 2013.
- [16] J. Yang, D. Zhang, J.-Y. Yang, and B. Niu, "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 650–664, 2007.
- [17] J. Chen, Z. Ma, and Y. Liu, "Local coordinates alignment with global preservation for dimensionality reduction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 1, pp. 106–117, 2013.
- [18] H. Li, X. Wang, J. Tang, and C. Zhao, "Combining global and local matching of multiple features for precise item image retrieval," *Multimedia Systems*, vol. 19, no. 1, pp. 37–49, 2013.
- [19] B. Yang and S. Chen, "Sample-dependent graph construction with application to dimensionality reduction," *Neurocomputing*, vol. 74, no. 1–3, pp. 301–314, 2010.
- [20] B. Yang and S. Chen, "Disguised discrimination of locality-based unsupervised dimensionality reduction," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 24, no. 7, pp. 1011–1025, 2010.
- [21] B. Yang, S. Chen, and X. Wu, "A structurally motivated framework for discriminant analysis," *PAA: Pattern Analysis and Applications*, vol. 14, no. 4, pp. 349–367, 2011.
- [22] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

- [23] X. Zhu, "Semi-supervised learning literature survey," Computer Sciences Technical Report 1530, University of Wisconsin-Madison, 2008.
- [24] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.
- [25] C. L. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*, Department of Information and Computer Sciences, University of California, Irvine, 1998.
- [26] D. Mateos-García, J. García-Gutiérrez, and J. C. Riquelme-Santos, "On the evolutionary optimization of k-NN by label-dependent feature weighting," *Pattern Recognition Letters*, vol. 3, no. 16, pp. 2232–2238, 2012.
- [27] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letters*, vol. 26, no. 2, pp. 181–191, 2005.
- [28] S. Pyatykh, J. Hesser, and L. Zheng, "Image noise level estimation by principal component analysis," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 687–699, 2013.
- [29] S. Liwicki, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Euler principal component analysis," *International Journal of Computer Vision*, vol. 101, no. 3, pp. 498–518, 2013.